

ABSTRACT

The present invention allows users to navigate in a Web application or Web pages using a combination of point-and-click and voice-input. At each point of the dialog, the user can use the standard point-and-click interface to perform context-dependent actions, or alternatively, use speech input to navigate and operate in the global application context. The voice input uses a voice navigation component which builds an interface to the installed recognition and synthesis engines. The point-and-click and the voice navigation components can be loaded automatically with the initial Web page of a Web application. Grammars for recognizing vocabulary related to that Web application will be provided with the voice navigation component. The present invention combines the advantages of a context-dependent point-and-click user interface with those of a context-independent speech-input interface. Accordingly, a multi-modal interface can be provided to a Web browser.